

Anastashia Pelletier
MATH 494
Assignment 3

1.

Tables:

Highest Team OPS+ by Year					
Year	Team	OPS+			
2000	SFN	1.215			
2001	SEA	1.200			
2002	NYA	1.162	2011	SLN	1.189
2003	BOS	1.188	2012	LAA	1.206
2004	CLE	1.156	2013	BOS	1.201
2005	NYA	1.174	2014	LAN	1.159
2006	NYA	1.135	2015	TOR	1.215
2007	NYA	1.174	2016	CHN	1.153
2008	TEX	1.150	2017	HOU	1.307
2009	NYA	1.174	2018	LAN	1.150
2010	CIN	1.133	2019	NYA	1.203

R Code:

```
library(tidyverse)
library(ggpubr)
library(gt)
library(readr)

#Importing Team file
teamsA3 <- read_csv("MATH 494/Teams (for
Assignment3)(in).csv")

#Creating subset for years 2000-2019
teamSub <- filter(teamsA3,
                  yearID %in% 2000:2019)
teamSub <- select(teamSub, yearID, teamID, franchID, G, W, L,
                  H, BB, HBP, AB, SF, `2B`, `3B`, HR, BPF)

#Get OBP and SLG for each year and by team
teamFun <- function(df){
  OBP <- with(df, round((H+BB+HBP)/(AB+BB+HBP+SF),3))
  TB <- with(df, (H+`2B`+2*`3B`+3*HR))
  SLG <- with(df, round(TB/AB,3))
}
```

```

df <- cbind(df, OBP, SLG)
}

teamSub <- teamFun(teamSub)

#Get Averages for OBP and SLG for each team
teamSub |> group_by(yearID) |>
  summarise(across(c(OBP, SLG), mean), .groups="keep") ->
  sumOPS
names(sumOPS)[2] <- "avgOBP"
names(sumOPS)[3] <- "avgSLG"
teamSub |>
  left_join(sumOPS |>
    select(yearID, avgOBP, avgSLG),
    join_by(yearID)) -> teamSub

#Get OPS+ for each year and by team
teamOPSF <- function(df){
  tOBP <- with(df, OBP/avgOBP)
  tSLG <- with(df, SLG/avgSLG)
  dOPS <- with(df, (tOBP+tSLG-1)/(BPF))
  OPSP <- with(df, round(dOPS*100, 3))
  df <- cbind(df, OPSP)
}

teamSub <- teamOPSF(teamSub)

#Create table for the team with the best OPS+ by year
tableSub <- select(teamSub, yearID, teamID, OPSP)

tableSub |>
  group_by(yearID) |>
  top_n(1, OPSP) |>
  ungroup(yearID)-> tableSub

tableSub <- gt(select(tableSub, yearID, teamID, OPSP),
  rowname_col = "yearID") |>
  tab_header(title = "Highest Team OPS+ by Year") |>
  tab_stubhead(label="Year") |>
  cols_label(teamID = "Team", OPSP = "OPS+")

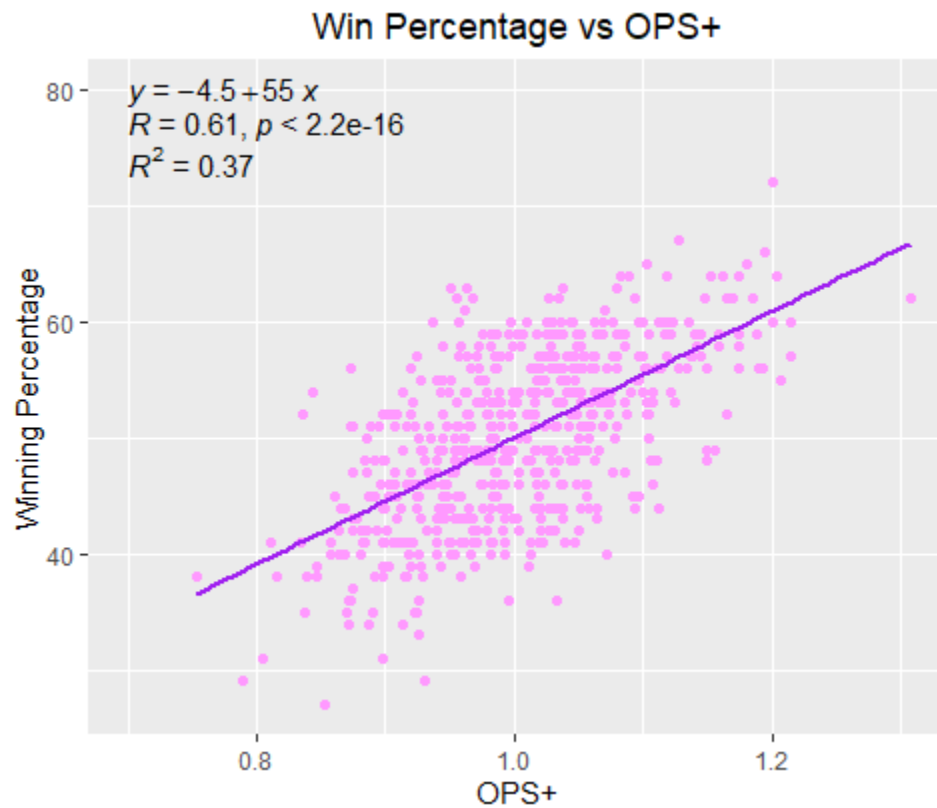
tableSub

#Export for Problem 2
write.csv(teamSub,
  "C:\\Users\\Anastashia Pelletier\\Documents\\MATH
  494\\teamSub.csv", row.names = TRUE)

```

2.

Scatterplot:



Analysis:

Since the R value is under 0.5 but above 0.3, we can say that there is a positive moderate correlation between the win percentage and the OPS+. Also, since the R^2 value is 0.13, we know that 13% of the variance in winning percentage is predicted by the team's OPS+ value.

R Code:

```
library(tidyverse)
library(ggpubr)
library(ggpmisc)
library(readr)

#Import Previous Dataset from Problem 1
teamSub <- read_csv("MATH 494/teamSub.csv")

teamWin <- select(teamSub, yearID, teamID, G, W, L, OPS+)

#Get Win Percentage
winF <- function(df){
  wPer <- with(df, round((W/(W+L))*100, 0))
```

```

df <- cbind(df, wPer)
}

teamWin <- winF(teamWin)

#Create Scatterplot for Win Percentage vs OPS+
Win_OPSP <- ggplot(teamWin, aes(OPSP, wPer)) +
  geom_point(color="#ff99ff") +
  geom_smooth(method = "lm", se=FALSE, color="purple") +
  stat_regline_equation(label.x = 0.7, label.y = 80) +
  stat_cor(label.x=0.7, label.y=77) +
  stat_cor(aes(label=..rr.label..), label.x=0.7, label.y=74) +
  labs(title="Win Percentage vs OPS+", x="OPS+", y="Winning
          Percentage") +
  theme(plot.title = element_text(hjust = 0.5))

Win_OPSP

#Create Linear Model for Win Percentage vs OPS+
model <- lm(wPer ~ OPSP, data=teamWin)
summary(model)

int <- round(coef(model)[1], 3)
slope <- round(coef(model)[2], 3)
r2 <- round(summary(model)$r.squared, 3)
r <- round(sqrt(r2), 3)

#Display Linear Model
glue::glue("y = {slope} + {int}x", "
          R = {r}", "
          R-Squared = {r2}")

#Exported For Fun
write.csv(teamWin,
          "C:\\Users\\Anastashia Pelletier\\Documents\\MATH
          494\\teamWin.csv", row.names = TRUE)

#PS, I can't resist using the color pink for my graphs...

```